基于多特征和深度神经网络的维吾尔文情感分类 *

买买提阿依甫，吾守尔·斯拉木，艾斯卡尔·艾木都拉†，杨文忠，帕丽旦·木合塔尔

(新疆大学 信息科学与工程学院，乌鲁木齐 830046)

摘　要：针对传统机器学习的情感分类方法存在长距离依赖问题，深度学习存在忽略情感词库的弊端，提出了一种基于注意力机制与双向长短记忆网络和卷积神经网络模型相结合的维吾尔文情感分类方法。将多特征拼接向量作为双向长短记忆网络的输入捕获文本上下文信息，使用注意力机制和卷积网络获取文本隐藏情感特征信息，有效增强了对文本情感语义的捕获能力。实验结果表明，该方法在二分类和五分类情感数据集上的 F1 值相比于机器学习方法分别提高了 5.59%、7.73%。

关键词：情感分类；双向长短记忆网络；卷积神经网络；注意力机制；维吾尔语

中图分类号：TP391.1　　doi: 10.19734/j.issn.1001-3695.2018.10.0809

Uyghur sentiment classification based on multi-features and deep neural network

Maimaitiayifu, Silamu Wushouer, Aimudoula Aisikaer†, Yang Wenzhong, Muhetaer Palidan

(*College of Information Science & Engineering*, *Xinjiang University*, *Urumqi* 830046, *China*)

Abstract: In order to solve the problem of long-distance dependence in traditional machine learning sentiment classification method and the disadvantage of ignoring the emotional lexicon in deep learning, this paper proposes a Uyghur sentiment classification method based on attention mechanism combined with bidirectional long-short term memory network and convolutional neural network model. The concatenated multi-feature vector is used as the input of the bidirectional long short-term memory network to capture the context information, the attention mechanism and convolution network are used to capture text hidden emotional feature information, which effectively enhances the capture ability of the text sentiment semantics. The experimental results show that the F1 value of this method on two-category and five-category Uyghur sentiment data sets higher than machine learning method 5.59%, 7.73%, respectively.

Key words: sentiment classification; bidirectional long short-term memory network; convolutional neural network; attention mechanism; Uyghur

0　引言

情感分析(sentiment analysis)，也称为观点挖掘，其目标在于挖掘文本中观点的态度、倾向和意见等主观感受[1,2]。人们一般以文本、音频和图像为载体，借助于一定的表情进行表达情感[3]。常见的情感分析研究主要包括情感识别、情感极性分析和主题情感分析[4]等。

随着信息技术的发展，大量基于维吾尔语的新闻网站、论坛、微博等平台也不断建立，促进了新疆地区教育和经济的发展，提升了新疆信息化水平[5]。与此同时，也有网民发布带有负面情感信息，对社会造成负面影响。通过情感分析技术对社交媒介的评论、舆论进行倾向性分析，能够帮助政府和安全部门及时了解网民的民意调查、舆论倾向及动态。因此，如何通过技术手段从文本中捕获用户的情感倾向信息是对新疆社会稳定和长治久安具有很重要的理论意义和应用价值。

国内外的学者已经对情感分析做了大量的研究，并取得了一些显著的成果。目前常见的情感分析方法主要分为基于词典的方法、基于传统机器学习的方法和基于深度学习的方法三种。基于词典的方法主要是构建包括情感词、情感短语的情感词典，其核心是词典和规则，通过点互信息量（pointwise mutual information, MPI）等方法[6]来计算词语的情感倾向，从而判断整个句子的情感极性。该方法受限于情感词典的覆盖度和判断规则的质量，而词典的构建和设计一个合适的判断规则一般都需要耗费很多人力和先验知识。基于传统机器学习的方法通常利用朴素贝叶斯(NB)[7]、最大熵（ME）[7]、支持向量机（SVM）[7]等进行情感分类。这些方法已被证明是简单有效的，但过度依赖背景知识和特征的选取，只能在有充足和正确标注的训练语料时才能够取得较高的分类效果，而这些高质量语料的标注和特征的选取仍然受到人为因素影响，需要投入大量人工成本。这些方法容易丢失文本深层语义信息，很难有效捕获文本里的情感信息。此外，此类方法对语料领域非常敏感，用一个领域的语料训练出来的模型不一定适应其他领域。

近几年来，随着深度神经网络技术的快速发展，深度神经网络针对资源丰富的英语或汉语等语言在机器翻译、语音识别、问答系统和文本摘要、关系提取和情感分析等各种 NLP 任务中取得了良好的性能。而由于维吾尔语资源匮乏，

收稿日期：2018-10-24；**修回日期：**2018-12-11　**基金项目：**国家自然科学基金资助项目（61363063）；国家“973”重点基础研究计划基金资助项目（2014CB340506）；新疆大学多语种重点实验室开放课题（XJDX0905-2013-01）

作者简介：买买提阿依甫（1981-），男（柯），博士研究生，主要研究方向为网络舆情分析关键技术；吾守尔·斯拉木（1942-），男（维），教授，院士，博导，主要研究方向为自然语言处理、语音信号处理、多语种信息处理；艾斯卡尔·艾木都拉（1972-），男（通信作者），教授，博士（后），主要研究方向为图像处理、语音信号处理（100425217@qq.com）；杨文忠（1971-），男，副教授，硕导，主要研究方向为舆情分析；帕丽旦·木合塔尔（1986-），女（柯），博士研究生，主要研究方向为语音合成技术.

情感分析研究较少。

维吾尔文情感分析工作起步较晚，目前还没有像英语或汉语一样丰富的情感资料可供使用。维吾尔文情感分析大多是基于情感词典和基于传统机器学习方法。采用深度神经网络进行情感分析的研究较少。

基于深度神经网络的情感分析方法主要是采用词向量对文本中的词语进行表示，进而构建句子、篇章的语义表示。在这些语义表示的基础上，采用深度神经网络模型对文本中蕴涵的情感信息进行学习，从而实现对文本情感的分析。目前常用于情感分析的神经网络模型包括循环神经网络(recurrent neural network, RNN)[8]、卷积神经网络(convolutional neural network, CNN)[9,10,21]、长短记忆网络(long short-term memory, LSTM)[11,12,20]、门控循环单元(gated recurrent unit, GRU)等[13,16]。

目前基于深度神经网络的情感分析方法大多是将文本看做整体进行语义表示，对于情感词或情感短语的体现没有突出，而基于情感词典的情感分析方法过度依赖情感词，没有考虑文本整体的语义关系。因此，为了解决上述问题，本文引入了注意力机制，采用注意力模型对维吾尔文本、情感词进行编码，提出了一种基于注意力机制的 BiLSTM-CNN 模型的维吾尔文情感分析方法。本文主要贡献如下：

a) 对现有的维吾尔文情感词典进行梳理，对“HowNet”和“NTUSD”汉文情感词库进行翻译并处理后构建了较完整的维吾尔文情感词库。

b) 本文使用音节特征向量、词性特征向量和位置特征向量，有效弥补了词向量的不足。

c) 提出了一种基于注意力机制的深度维吾尔文本情感分类方法（ATT-BiLSTM-CNN），通过注意力模型使模型获取更深层次的情感特征信息，有效增强了对文本情感语义的捕获能力，从而提高了最好的情感分类效果。

d) 对比了本文模型和基准模型在情感二分类和五分类数据集上的分类效果，验证了本文模型的有效性。

1　相关工作

1.1　注意力机制

注意力机制模仿人脑关注事物的某关键部分分配更多的注意力，通过计算注意力概率分布，将最关键的重要部分更加突出，从而对传统深度学习模型起到优化作用。2014 年 Mnih 等人[14]在图像分类任务中第一次引入了注意力机制，达到了较好的分类效果，验证了注意力机制在图像分类任务中的有效性。随后，Bahdanau 等人[15]借鉴注意力机制在图像分类任务中的应用，将注意力机制引入到机器翻译任务中，使注意力机制成为自然语言处理领域的热点。随着研究的推进，各种改进的带注意力机制模型在文本摘要生成、文本分类、句法分析、情感分类、短文本对话等任务中均取得了良好的效果。

随着深度神经网络和注意力模型在自然语言处理领域的广泛应用，基于注意力机制的各种改进模型也应运而生，因为深度神经网络能够有效学习到文本特征信息，有效解决特征表示问题，从而提高了情感分类等 NLP 任务的准确率。Tang 等人[16]提出了两种带主体信息的基于注意力机制的 TD-LSTM 和 TC-LSTM 两种模型，提高了情感分类的准确率。Luong 等人[17]提出了一种局部注意力模型，通过调整窗口长度，在指定的窗口范围内计算每个单词的对其概率。Wu 等人[18]提出了基于注意力机制的 CNN-LSTM 模型，在主题情感分析任务中达到了最好的性能。2015 年 Yin 等人[19]提出一种融合注意力机制的 CNN 网络。这些方法的提出验证了注意力机制和深度神经网络结合的有效性。

受文献[18,19]的启发，本文提出了一种基于注意力机制的 BiLSTM-CNN 模型，使用 BiLSTM 和 CNN 模型分别捕获了文本历史、未来上下文信息和局部信息，并通过注意力机制使模型更加关注文本中的带情感信息的部分，从而提高了维吾尔文情感分类的准确率。

1.2　神经网络

由于深度神经网络相对于传统的机器学习方法拥有优秀的自学习特征提取的能力，深度学习模型在自然语言处理的各领域得到了广泛应用。CNN 最初由 LeCun 等人[21]提出用于计算机视觉，在计算机视觉任务中得到了很好的效果。前人已经证明 CNN 在潜在特征展示方面表现出色。例如词性标注、句子分类等。2014 年 Kim 等人[9]使用 CNN 解决了情感分类问题，并通过实验证明了 CNN 的分类性能优于递归神经网络。Kalchbrenner 等人[28]提出了一种采用 *k*-max 池化和多层卷积神经网络相结合的新颖模型。另一个流行的深度学习模型是序列模型，即循环神经网络（RNN），它通过隐藏状态能够保留文本的历史信息，因此，它更好地捕获文本中词语之间的语义关系。RNN 的变体模型已经成功应用于机器翻译、文本生成等任务。LSTM 可以捕获到文本中的长依赖关系，能够从整体上理解评论信息的情感语义。

近几年来，由于 CNN 和 LSTM 的建模能力是相辅相成的，研究人员一直致力于将它们结合起来。目前未有见到融合注意力机制和深度神经网络的方法来解决维吾尔文情感分类的报道。

2　方法

在本章中将详细介绍维吾尔文句子情感分类的基于注意力机制的 BiLSTM-CNN 神经网络 （AT-BiLSTM-CNN）的体系结构。本文所提出的模型主要由词向量表示层、BiLSTM 层、注意力层、CNN 层和情感计算层组成。每个层次的输出作为下一个网络层的输入，其中词向量表示层中词向量、词性向量、音节向量和位置向量拼接的向量作为 BiLSTM 网络层的输入进行训练，BiLSTM 模型的输出作为注意力层的输入进行编码，注意力层的输出作为 CNN 网络的输入进行进一步特征提出训练，最后通过情感计算层输出情感分类结果。基于注意力机制的 BiLSTM-CNN 网络模型框架如图 1 所示。

2.1　词向量表示层

本文模型框架最底层是词向量表示层，是整个框架的输入。假设由 n 个单词组成的句子 $S=\{s_1, s_2, \ldots, s_n\}$，每一条维吾尔文句子由词向量矩阵 R_W、词性向量矩阵 R_{POS}、音节向量矩阵 R_{SYL} 和位置向量矩阵 R_{LOC} 构成的句子向量矩阵 $R_S=[R_w^T \oplus R_{pos}^T \oplus R_{syl}^T \oplus R_{loc}^T]$，其中 $\oplus$ 是向量拼接操作。

2.1.1 词向量

词向量表示（word embedding）[22,23]基本思想是将词语转换为计算机能够处理的低维度，稠密实数向量，将词语间的语义关系较好映射到词向量间的位置关系。词向量具有很好的计算特性，即使对词向量进行简单数学运算后，仍然可以保留文本特征信息，从而在很大程度上缓解了维度灾难。为了生成词向量查询表，本文使用对大规模无标记语料进行训练生成词向量。本文实验中利用 Python 库中 Gensim 工具的 skip-gram 模型，用大规模句子语料训练出了维吾尔文词向量模型。对于给定的语料库，本文将单词向量存储在一个词向量查询矩阵 $M=R^{|V|\times d_w}$。其中：$|V|$是给定无标记维吾尔文

本语料库的词汇量；d_w 是单词向量的维数。对于句子 $S=\{s_1, s_2, \ldots, s_n\}$，$n$ 是句子长度。句子矩阵表示为 $X=\{x_1, x_2, \ldots, x_n\}$，$x_i$ 是单词 s_i 从词向量查询矩阵 M 中获取的词向量。

假设语料库由 $w_1,w_2,\ldots,w_m$ 个单词组成，Skip-Gram 模型的目标是使以下函数最大化：

$$F=\frac{1}{M}\sum_{m=1}^{M}\sum_{-n\le j\le n, j\ne 0}\log p(w_{t+j}\mid w) \tag{1}$$

其中：n 是训练窗口大小的参数。本文实验在词向量模型训练中，上下文窗口大小设置为 5，迭代次数为 8 词，分别生成了维度为 100、200、300 和 400 的词向量模型。对于未登录词，本文采用均匀分布 $U(-0.01,0.01)$来随机初始化词向量。

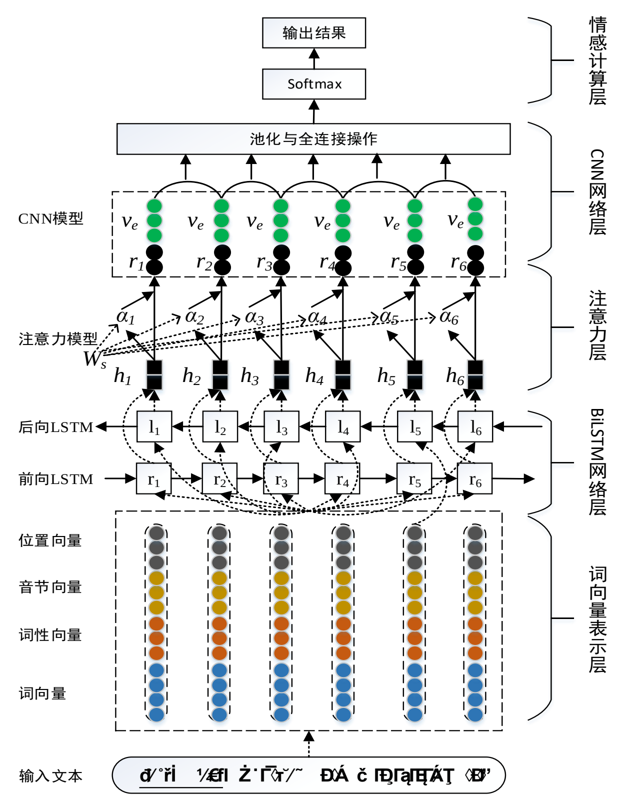


图 1　基于注意力机制的 BiLSTM-CNN 网络模型框架

Fig. 1　Bilstm-CNN model based on attention mechanism

2.1.2 词性向量

词性特征包含了词语丰富的语义信息，将词性特征转换成词性向量后，可以将其作为神经网络模型的输入，能够进一步发现语句中词语的结构联系以及单词情感信息。例如“بەك”或者“ئاراق”等形容词可以用来表示情感强度，神经网络模型通过词性特征可以发现这样的情感词语，因此本文在本文中加入了词性特征，进一步提高了情感分类准确率。

目前，维吾尔文没有一个统一规范的词性标注集。新疆多语种信息技术实验室等各研究单位都建立了自己的词性标注集。新疆多语种信息技术实验室手工建立了包含 120 万个单词的词性标注集，其中包括一级词性标注集(15 个标签)（表 1）、二级词性标注集（71 个标签）、三级词性标注集（51 个标签）。

为了更好地学习句子中情感词的情感特征信息，本文在一级词性标注集的基础上增加了两个词性，即正面情感词重新标注为“POS”标签、负面情感词重新标注为“NEG”标签，使情感词的词性更加突出。实验使用 17 位的 one-hot 向量表示词性向量（以下 v_{pos} 表示词性特征向量，R_{pos} 表示词性向量矩阵），然后与词向量、音节向量和位置向量拼接生成混合向量作为 BiLSTM 模型的输入，提高了情感分类的准确率。

表 1　维吾尔语词性一级标记集

Table 1　Uyghur first level part of speech tag set

序号	名称	标记	序号	名称	标记
1	名词	N	10	语气词	T
2	动词	V	11	标点符号	Y
3	形容词	A	12	附加成分	X
4	副词	D	13	后置词	R
5	代词	P	14	拉丁文	L
6	数词	M	15	量词	Q
7	连词	C	16	正面情感词	POS
8	模拟词	I	17	负面情感词	NEG
9	感叹词	E			

2.1.3 音节向量

维吾尔语中的否定词和程度词对情感词的极性和强度有很大的影响，否定词会改变情感词极性，程度词会增强或削弱句子情感强度。维吾尔语的否定句以动词的否定形式来表示，否定句中的动词作为谓语出现在句末，而句子中的否定意义通过句末尾动词加否定后缀，也就是加[ما，مە]等否定音节来表达。譬如：“قىل + ما = قىلما”。同样，维吾尔语中的程度副词，通过名词和形容词词尾加表示程度的后缀，而表达句子的副词程度，构造方法是名词和形容词词尾加[رەك，راق]等音节，譬如：“سەت + رەك = سەترەك”。维吾尔语中表达否定和程度副词的方法，只要通过动词、名词和形容词词尾加表示否定和程度的音节来完成的。因此，在本文中，对给定的维吾尔情感句子，先分词再分音节，找出最关键的决定情感的最小单位进行了更细更深得研究。最终，将音节向量与词向量和词性向量拼接构成的混合向量作为模型的输入进行训练，提高了维吾尔文情感分类的准确率。维吾尔文单词可以由多个音节也可以由一个音节构成，本文可以用以下形式：

单词={音节 $_n$，音节 $_{n-1}$，…，音节 $_2$，音节 $_1$}

切分单词音节时本文使用了新疆多语种实验室开发的音节切分工具。本次实验中采用了从后向前切分的方式，也就是从一个单词的尾部音节依次提取音节。比如，给定的单词(学生们的)ۋوقۇغۇچىلارنىڭ，它的音节表示形式为 ۋوقۇغۇچىلارنىڭ=ۋوقۇ+غۇ+چى+لار+نىڭ,则 $F_{syl}=[F_{syl4}; F_{syl3}; F_{syl2}; F_{syl1}]$,其中：$F_{syl}$= [نىڭ]，$F_{syl2}$=[لار]，$F_{syl3}$=[چى]，$F_{syl4}$=[ۋوقۇ]。

本次实验中，本文使用均匀分布 $U(-0.01,0.01)$来随机生成维度为 30 的音节向量。

2.1.4 情感词位置向量

本文对实验室提供的维吾尔文情感语料库进行详细统计与分析，发现具有情感的感叹词和模拟词一般出现在句首，维吾尔语中感叹词和模拟词一般出现在句首表示句子的情感强劲，后用逗号隔开，感情强烈时，可用感叹号表示。出现在句末的情感词表示肯定或否定，情感词的词性一般为动词。在句中和句末倒数第二个位置出现的词性情感词一般为形容词和程度副词。维吾尔文情感词在句子中的位置示例如表 2 所示。

表 2　维吾尔文情感词在句子中的位置示例

Table 2　Examples of location of Uyghur sentimental words in sentences

序号	示例	翻译	情感词位置
1	قىزىۋاتىدۇ ئۇسسۇل تازا بىلەن ناخشا	歌舞太热闹了	句末
2	پاھ ، بۇ گۈللەرگە قارا پاھ	哇哇，看这些花儿	句首
3	بولاتتى ياخشى ياردەملەشسىڭىز ماڭا سىز	要是您帮我就好了	句末
4	ئىكەن چىرايلىق ھەققەتەن بەك يەر بۇ	这个地方太美丽了	句末倒数第二

为了提高维吾尔情感分类的准确率，本文探索获取更多的情感信息作为特征，从而提高分类准确率。对实验室提供的 6 万多条维吾尔文情感句子中的情感词的位置进行了统计，大部分情感词出现在句首和句末（表 3）。

表 3　维吾尔情感词在居中的位置统计

Table 3　Statistics of locations of Uygur sentimental words in sentence

序号	位置	百分比(%)
1	句首	21
2	句末	45.7
3	句末倒数第二位置	19
4	句中	14.3

在训练模型前，本文对句子中的每个单词使用 one-hot 表示进行向量化，如果当前单词是情感词，对应的位置设为 1，否则为 0。本文将情感语料库中最长的句子长度 23 作为位置向量维度（以下用 v_{loc} 来表示位置向量，R_{loc} 表示由位置向量组成的矩阵）。最后，将词向量、词性向量、音节向量和位置向量拼接生成的混合向量作为模型的输入进行训练。

2.2　BiLSTM 网络层

2.2.1 长短期记忆单元 LSTM

循环神经网络(recurrent neural networks, RNN)[11]作为深度学习领域比较常用的序列学习方法，通过循环递归的结构使得自身具有一定的记忆能力，理论上可以获取任意长度历史信息，然而在实际应用中经常出现梯度消失或梯度爆炸问题，从而无法学习输入序列中距离较远的逻辑关系。

为了解决 RNN 模型的不足，Hochreiter 等人[20]提出了 LSTM 模型，替换 RNN 模型中的隐含层，从而避免了 RNN 模型的梯度消失问题。一个 LSTM 模型包含输入门 i_t、输出门 o_t、遗忘门 f_t 和记忆单元 c_t，其中记忆单元是核心部分，如图 2 所示。对于一个由 n 个单词构成的维吾尔文句子的词向量序列 $S=\{x_1, x_2, \ldots, x_n\}$，$x_t$ 为 LSTM 单元在第 t 步的输入，表示输入序列中单词对应的混合向量。

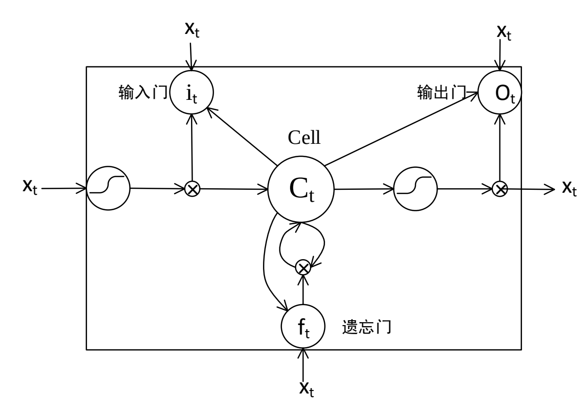


图 2　LSTM 单元结构

Fig. 2　Structure of LSTM unit

LSTM 单元中的三个门和记忆单元的计算如下：

$$X = \begin{bmatrix} h_{t-1} \\ x_t \end{bmatrix} \tag{2}$$

$$f_t = \sigma(W_f \cdot X + b_f) \tag{3}$$

$$i_t = \sigma(W_i \cdot X + b_i) \tag{4}$$

$$o_t = \sigma(W_o \cdot X + b_o) \tag{5}$$

$$c_t = f_t \odot c_{t-1} + i_t \odot \tanh(W_c \cdot X + b_c) \tag{6}$$

$$h_t = o_t \odot \tanh(c_t) \tag{7}$$

在上述计算公式中：$W_f, W_i, W_o \in \mathbb{R}^{m \times d}$ 表示权重矩阵；$b_f, b_i, b_o \in \mathbb{R}^m$ 表示偏置；σ 是非线性激活函数；m=100 是 LSTM 网络的单元数量；⊙表示元素逐个点乘积；xt 包含 LSTM 单元的输入向量；$h_t \in \mathbb{R}^m$ 是隐藏层向量。对于一个给定的句子 S={ x1, x2, …, xn }，每个单词 xi 将映射到对应的词向量 $v_w^i \in \mathbb{R}^{dw}$，每个单词的词性映射到词性向量 $v_{pos}^i \in \mathbb{R}^{dp}$,单词的音节向量映射到音节向量 $v_{syl}^i \in \mathbb{R}^{ds}$，提取当前单词的音节向量根据 3.1.3 节中介绍的音节切分方法，单词的位置向量映射到位置向量 $v_{loc}^i \in \mathbb{R}^{dl}$（上述中 dw=300、dp=17、ds=30、dl=23 分别为词向量、词性向量、音节向量和位置向量的维度，因此在神经网络的输入层混合词向量的维度为 d=dw+dp+ds*n+dl，其中 n 是从单词提取到的音节特征数量，训练前人工设定），最后拼接这些向量后生成句子向量矩阵，即

$$[x_1, x_2, \ldots, x_n] \Rightarrow [R_w^T \oplus R_{pos}^T \oplus R_{syl}^T \oplus R_{loc}^T] \Rightarrow \begin{bmatrix} v_w^1 \oplus v_{pos}^1 \oplus v_{syl}^1 \oplus v_{loc}^1 \\ v_w^2 \oplus v_{pos}^2 \oplus v_{syl}^2 \oplus v_{loc}^2 \\ \vdots \\ v_w^n \oplus v_{pos}^n \oplus v_{syl}^n \oplus v_{loc}^n \end{bmatrix} \tag{8}$$

2.2.2 双向长短期记忆网络

LSTM 能够捕获输入序列的长期历史信息，但无法捕获未来信息。双向长短期记忆网络（BiLSTM）是由向前 LSTM 网络与向后 LSTM 网络叠加构成的网络，它使用向前和向后 LSTM 网络能够分别获取历史信息和未来信息，从而获取更多上下文依赖关系。向前和向后 LSTM 的两个隐藏状态 $\overrightarrow{h_t}$ 和 $\overleftarrow{h_t}$ 分别是关于过去和未来的信息。由它们拼接生成的向量 $h_t = [\overrightarrow{h_t}, \overleftarrow{h_t}]$ 提供了完整的上下文历史与未来信息，向前向后的输出最后是一个融合的结果 ht。句子中每个单词都以词向量的形式嵌入网络，通过使用 BiLSTM 对维吾尔文本进行编码，实现前向的语义信息和后向的语义信息以相同的地位被考虑，从而获取文本中的情感信息。

2.3　注意力层

众所周知，根据句子的不同部分，句子的情感通常会有所不同。句子中的某些情感词或短语对于句子的情感有决定性作用，而其他单词则无关紧要。因此，本文引入了注意力机制来关注这些重要词语，并将它们的表示形式转换为句子向量。实际上，注意力机制是计算句子中的上下文向量。

本文模型中，将由词向量、词性向量、音节向量和位置向量拼接构成的词向量作为 BiLSTM 网络的输入进行编码，将 BiLSTM 网络在每个时间步骤产生的一个隐藏状态向量 h_t 输入到注意力模型中，采用注意力模型对维吾尔文句子进行编码。注意力模型首先使用单层感知器（MLP）对输入隐藏状态 h_t 进行加权，从而为输入的句子构造一个上下文相关的句子文本表示向量 r，如式（9）所示。

$$r_t = \alpha_t h_t \tag{9}$$

其中：α_t 为 BiLSTM 网络输出的隐藏状态向量 h_t 的注意力权重，它与每一个时间步骤的输入状态和一个随机初始化的上下文向量 m_w 有关，α_t 可以通过式（10）（11）计算得到。

$$\alpha_t = \frac{\exp(m_t^T m_w)}{\sum_t \exp(m_t^T m_w)} \tag{10}$$

$$m_t = tanh(W_w h_t + b_w) \tag{11}$$

式（11）中：W_w 为模型中的权重；b_w 为模型中偏置，它们与 m_w 一起作为模型的参数，通过不断训练学习得到。通过以上的注意力模型，可以为模型的输入构造一个定长的上下文相关的文本表示向量 r，此向量中的信息包含各个输入状态的重要程度，使模型更加关注文本里带情感信息的情感词或短语，从而提高维吾尔文情感分类的准确率。通过这样的文本表示，可以为特征序列分配不同的注意权重。最后，可以更容易地识别诸如情感词之类的重要信息。

2.4　CNN 网络层

卷积神经网络(convolutional neural network, CNN)[9,10,21]能够捕获文本中的局部特征信息，在句子级别的情感分类任务中表现优秀[9]。因此，在本文中本文使用 CNN 网络作为 BiLSTM 模型的补充，实现了带注意力机制的 BiLSTM-CNN 混合模型（简称为 AT-BiLSTM-CNN）。

卷积神经网络由输入层、卷积层、池化层和全连接层组成。本次实验中将注意力模型输出的 r_t 与情感词向量 v_e 拼接的向量矩阵 r^*作为输入进行训练。即

$$r^*=[r_t \oplus v_e] \tag{12}$$

卷积层使用不同的录波器对输入向量进行卷积操作，从而获取输入向量中的局部特征。其计算公式如下：

$$u(i) = f(W \bullet r^*_{i:(i+h-1)} + b) \tag{13}$$

其中：$r^*_{i:(i+h-1)}$ 表示输入向量的第 i 行至 $i+h$-1 行抽取的局部特征矩阵；$u(i)$是某个卷积核在位置 i 的卷积输出；W 是滤波器；b 是卷积偏置项；$f(\cdot)$是非线性卷积核函数，本文用 relu 作为激活函数。

之后，最大池化层对向量 u 中所有 $u(i)$求最大值以获取其中最显著的特征值，计算公式如下：

$$\hat{c}_i = \max_i \{u_j(i)\} \tag{14}$$

其中：j 表示第 j 个卷积核。本文实验中使用了不同尺寸的卷积核，分别为 h=3，h=4，h=5。对于有 T 个卷积核的窗口采样得到的特征信息如式（15）所示。下采样层输出的特征向量作为全连接层的输入。

$$\hat{c} = [\hat{c}_1, \hat{c}_2, \ldots, \hat{c}_T] \tag{15}$$

2.5　情感计算层

本文中使用 *softmax* 函数[24,25]作为情感分类器，将全连接层输出的特征表示 $\hat{c}$ 作为 *softmax* 函数的输出情感极性分类判别的结果，计算公式如下：

$$\hat{y} = soft\max(W_f \bullet \hat{c} + b_f) \tag{16}$$

其中：$\hat{y}$ 是模型预测的文本情感类别；W_f、b_f 分别是全连接层权重矩阵和偏置。

2.6　模型训练

本文使用反向传播算法来训练和更新模型，通过最小化交叉熵来优化模型[26,27]。交叉熵损失函数的计算公式如下：

$$loss = -\sum_{i=1}^{T}\sum_{j=1}^{C} y_i \log \hat{y}_i^j + \lambda \| \theta \|^2 \tag{17}$$

其中：T 是训练数据集；C 为情感类别数；y 为文本实际情感类别；λ 为正则项，是损失函数的惩罚项；θ 是设置的参数。

3　实验准备

3.1　情感词库的构建

情感词库包括情感词、情感短语、跟情感词搭配的否定词、副词、形容词、感叹词、语气词以及一些习惯用语构成的具有情感倾向的单词或短语。

目前尚未有公开的维吾尔文情感词库。对于资源匮乏的语言，有研究通过尝试将资源丰富的语言的情感词库翻译成资源稀缺的语言，并取得较好的分类效果。因此，本文首先使用新疆信息技术实验室开发的汉维翻译接口对目前使用率比较普遍的“HowNet”汉文情感分析词语表中的正面情感、负面情感词语、正面评价词语和负面评价词语四个情感词语表以及　　大学创建的“NTUSD”汉文情感词典进行翻译；然后人工对齐进行整理，去除部分翻译后失去情感倾向的单词和短语；除此之外，本文人工收集情感句子中的情感词；最后构建了包括 5 643 个词汇的情感词典（UySentiDict），其中正面情感词语有 2 411 个，负面情感词语有 3 232 个。

本文实验中将句子中的情感词转换为词向量，并与注意力层输出的隐藏状态向量拼接生成的向量作为 CNN 网络的输入，从而起高了情感分类准确率。

3.2　情感分析数据集

为了构建维吾尔文句子级情感语料，本文实验选取天山网等维文网站上公开发布的文章和评论信息。本文利用实验室开发的网络爬虫工具下载网页，经过去重、去噪等操作筛选出包含情感倾向的评论信息。对收集好的语料进行分句，然后人工对其进行分类，构建了二分类（UySenti2Data），如表 4 所示。

表 4　维吾尔二分类数据集

Table 4　Uyghur two-group sentimental data set

数据	正面句子	负面句子	总计
训练集	1.2 万	1.2 万	2.4 万
开发集	0.15 万	0.15 万	0.3 万
测试集	0.1 万	0.1 万	0.2 万
总计	1.45 万	1.45 万	2.9 万

除了维吾尔文情感二分类（正面、负面）语料库外，本文还构建了情感五分类（中性、高兴、生气、惊讶、难过）语料库（UySenti5Data），如表 5 所示。

表 5　维吾尔五分类数据集

Table 5　Uyghur five-group sentimental data set

数据	中性	高兴	生气	惊讶	难过
训练集	1.5 万	1.2 万	1.2 万	0.8 万	0.8 万
开发集	0.15 万	0.1 万	0.1 万	0.1 万	0.1 万
测试集	0.1 万	0.1 万	0.1 万	0.1 万	0.1 万
总计	1.75 万	1.4 万	1.4 万	1 万	1 万

为了验证本文混合模型的有效性，在情感二分类数据集的基础上，构建了情感五分类数据集，表 6 中给出了每个类别的举例和对应的翻译。

表 6　维吾尔五分类数据举例说明

Table 6　Examples of Uyghur five-group sentimental data set

类型	维文情感语句举例	翻译
中性	بىز تەجربە ئىشلەۋاتىمز	本文在做实验
高兴	ئانا ۋەتىنىمزنى قىزغىن سۆيىمز!	本文热爱祖国！
生气	سەن نىمانداق گەپ ئاڭلىمايسەن؟	你怎么不听话？
惊讶	ۋاھ ، ھەقىقەتەن مۆجىزە يۈز بەردى	哇，真的发生了奇迹
难过	ۋاي ئىسست دوختۇرخانىغا ئۈلگۈرمىدى	哎，很遗憾，没赶上医院

3.3　评价指标

文本情感分类作为一种文本分类任务，为了评估本文模型的情感分类性能，本文使用准确率（precision）、召回率（recall）和 F 值作为评价标准。计算情感分类的准确率、召回率和 F1 值公式如下：

$$\begin{cases} P = \dfrac{TP}{TP+FP} \\ R = \dfrac{TP}{TP+FN} \\ F1 = \dfrac{2 \times P \times R}{P+R} \end{cases} \quad (18)$$

其中：TP（true positive）是将正类预测为正类的句子数；FN（false negative）是将正类预测为负类的句子数；FP 是将负类预测为正类的句子数；TN 是将负类预测为负类的句子数。

3.4　数据预处理

由于从网上下载的文本语料不是规范正式，这种非正式语料通常包含 URL、邮箱、各种标点符号、数字、用户名等，而这些非正式文本给情感分类带来了极大挑战。

为了去除噪声干扰，本文预先对维吾尔文情感文本语料进行预处理。其具体规则如下：

a) 所有数字都被替换为“0”处理。

b) 如果句子里存在网址和电子邮件字符串，则本文对其进行删除。

c) 删除除了句号、逗号、感叹号、问号以外的标点符号，以移除任何可能影响本文情感分类性能的噪声。

d) 如果句子中存在汉文单词则用实验室提供的汉维翻译接口对齐进行翻译，用翻译结果替换源汉文单词。

e) 如果情感句子中出现连续的感叹号、问号、句号或其他字符，则只保留一个对应的符号或字符。

3.5　实验参数设置

为了达到理想的情感分类效果，本文进行了反复实验，调节了模型的超参数。通过交最小化交叉熵证来选择实验参数。单词向量维度设置为 300，LSTM 单元的大小设置为 100。dropout 设置为 0.5，以防止过度拟合。使用样本数量(mini-batch)为 128 和 Adam 优化算法来训练模型。在实验中，模型训练的迭代的次数为 19 时达到了最好的准确率，更多细节如表 7 所示。

表 7　神经网络参数设置

Table 7　Neural network parameter setting

网络层	超参数	取值
LSTM	LSTM 单元数	100
CNN	滤波器窗口大小	3,4,5
	滤波器数量	100
文本表示层	词向量	300
	音节向量	30
	词性向量	17
	位置向量	23
注意力层	注意力隐藏层数量	100
其他参数	dropout	0.5
	优化函数	Adam
	样本数量(batch size)	32
	初始学习率(learning rate)	0.01
	网络迭代次数(epoch)	19

3.6　实验对比模型

本文使用不同的情感分类方法与本文提出的基于多特征和注意力机制的 BiLSTM-CNN 混合模型进行对比，验证本文提出的方法的有效性，并使用词向量、词性向量、音节向量和位置向量特征提高了本文方法的情感分类准确率。

a) 基于情感词典的方法（SentiDict）:根据实验室构建的情感词典来对维吾尔文进行情感分类，如果句子包含正面情感词就分为正面、如果包含负面情感词就分为负面情感句子。

b) 多项式朴素贝叶斯 (multinomial naive Bayes，MNB)：是一个典型的传统机器学习方法之一，在许多文本分类和情感分类任务中得到了广泛应用。

c) CNN 模型：采用单通道和多通道卷积神经网络对维吾尔文句子的词向量表示进行学习，通过卷积和最大池化操作捕获维吾尔文句子的局部特征实现文本的情感分类。

d) 支持向量机（support vector machine，SVM）模型：SVM 是最常用的传统机器学习情感分类方法，通常采用 n-gram 作为特征进行分类，本次实验中本文使用 unigram、bigram 和 trigram 作为 SVM 的特征进行分类。

e) 长短记忆网络（LSTM）：使用标准的单向 LSTM 网络对维吾尔文句子进行编码表示映射到可变长目标序列。

f) 双向长短记忆网络（BiLSTM）:不带注意力机制的 BILSTM 模型对维吾尔文句子进行情感分类。

g) BiLSTM-CNN（BiLSTM）:不带注意力机制的 BiLSTM-CNN 混合模型对维吾尔文句子进行情感分类。

h) 基于注意力机制的 BiLSTM-CNN（ATT-BiRNN-CNN）模型：本文提出的引入注意力机制的 BiLSTM-CNN 混合模型。

4　实验结果及分析

为了验证本文提出的 **ATT-BiRNN-CNN** 模型在维吾尔文情感分类任务上的有效性，在实验室构建的二分类和五分类情感数据集上进行对比实验。两个情感分类数据集中分配的训练数据集、开发集和测试集的具体数据量如表 4、5 所示。

本文实验中选取了基于词典的分类方法、典型的传统机器学习和简单神经网络模型与本文提出的模型进行了对比。

4.1　神经网络方法与基准方法对比实验

为了对比神经网络方法与词典方法和传统机器学习方法，第一组实验在 UySenti2Data 和 UySenti5Data 两种情感分类数据集上进行了实验。为了公平对比各方法的分类效果，除了词典方法外，其他机器学习方法和神经网络方法都只使用本文预先训练出来的维度为 300 的词向量作为输入进行实验，如表 8 所示。

表 8　基准模型对比实验结果

Table 8　Benchmark model comparison experiment results

模型	情感 2 分类			情感 5 分类		
	准确率/%	召回率/%	F1 值	准确率	召回率	F1 值
词典方法	56.23	61.89	54.44	44.5	46.72	43.07
MNB	62.54	61.34	61.93	52.2	50.86	51.52
SVM-unigram	77.4	78.95	78.12	64.5	63.78	63.14
SVM-bigram	84.23	82.87	83.54	72.71	71.83	72.27
SVM-trigram	81.5	80.24	80.87	71.65	69.24	70.42
LSTM	72.59	71.2	71.89	59.73	58.48	59.1
CNN	83.65	82.64	83.14	74.51	74.24	74.37
BiLSTM	74.23	73.34	73.78	68.82	67.32	68.06
BiLSTM-CNN	85.47	84.57	85.02	74.57	73.29	73.92
ATT-BiLSTM-CNN	**86.28**	**85.89**	**86.08**	**76.17**	**75.83**	**76.0**

从表 8 可以看出，本组实验中对词典、MNB、SVM 机器学习方法，以及 LSTM、CNN、BiLSTM、BiLSTM-CNN、ATT-BiLSTM-CNN 模型进行了分类，实验结果表明，本文提出的 ATT-BiLSTM-CNN 模型的二分类、五分类情感 F1 值均高于其他基准方法。

在实验过程发现，针对词典方法，如果情感词不在情感词典中就无法对句子进行正确分类。MNB 需要大量的语料

才能达到较好的分类效果，SVM 方法具有较强的泛化能力，在 SVM 几个模型中 SVM- bigram 模型的表现最佳。单独深度神经网络模型中 CNN 模型的分类性能较好，它可以获取情感词的局部特征来提高分类效果。混合模型中本文提出的混合模型的分类效果最佳。表 9 是各模型对同一个语句 S=“باشقىلارنىڭ نەرسىسىنى رۇخسەتسىز ئالما”（ 未经他人允许不要拿别人的东西）的情感五分类结果.

表 9　各模型对同一个实例的分类结果

Table 9　Sentiment classification results of each model for same instance

模型	情感 2 分类	情感 5 分类
词典方法	正面	中性
MNB	正面	中性
SVM-unigram	正面	中性
SVM-bigram	**负面**	**生气**
SVM-trigram	正面	难过
LSTM	正面	难过
CNN	**负面**	**生气**
BiLSTM	负面	惊讶
BiLSTM-CNN	**负面**	生气
ATT-BiLSTM-CNN	**负面**	**生气**

从表 9 中可以看出，SVM-bigram、CNN、BiLSTM-CNN 模型和 ATT-BiLSTM-CNN 对实例的情感二分类和五分类结果是正确的。

4.2　多特征对情感分类的影响

为了验证词性特征、音节特征和位置特征对情感分类的有效性，本组实验中采用不同的特征组合作为模型输入。本组实验中依次将词性向量（F_{pos}）、音节向量（F_{syl}）和位置向量（F_{loc}）添加到词向量（F_w）拼接生成的混合向量（$F= F_{pos} \oplus F_w \oplus F_{syl} \oplus F_{loc}$）作为模型输入进行训练。首先，本文将词性向量与词向量拼接生成的向量（$F= F_{pos} \oplus F_w$）作为模型的输入进行训练，其二分类和五分类 F1 值分别提升了 0.26%、0.96%。其次，在词向量和词性向量的基础上增加了音节向量，其中 F_{syl1} 是单词最后一个音节，F_{syl2} 是单词最后两个音节，F_{syl3} 是单词最后三个音节，其他依此类推，对于音节数不够的单词进行补 0 的方式进处理。本次实验中尝试最多截取单词最后 4 个音节作为特征进行测试。从表 8 可以看出，F_{syl3} 作为特征时二分类和五分类准确率均得到了明显的提升，分别提升了 1.57%、2.19%。最后，增加了位置向量作为词向量的补充进行训练，二分类和五分类 F1 值分别提升了 3.05%、4.0%。本组实验结果如表10所示。

表10　多特征对比实验

Table 10　Benchmark model comparison experiment results

模型及特征向量		情感 2 分类			情感 5 分类		
模型	特征	准确率/%	召回率/%	F1 值	准确率	召回率	F1 值
ATT-BiLSTM-CNN	$F_w \oplus F_{pos}$	87.67	86.34	87.0	77.68	76.25	76.96
	$F_w \oplus F_{pos} \oplus F_{syl1}$	87.78	87.12	87.45	77.9	76.47	77.18
	$F_w \oplus F_{pos} \oplus F_{syl2}$	88.1	87.71	87.90	77.43	76.63	77.03
	$F_w \oplus F_{pos} \oplus F_{syl3}$	**88.91**	**88.23**	**88.57**	**79.49**	**78.82**	**79.15**
	$F_w \oplus F_{pos} \oplus F_{syl4}$	87.82	87.24	87.53	78.67	77.29	77.97
	$F_w \oplus F_{pos} \oplus F_{syl3} \oplus F_{loc}$	**89.76**	**88.51**	**89.13**	**80.26**	**79.76**	**80.0**

维吾尔语中兼类词和未登录词数量较多，在文本中出现次数的频繁率也高，传统统计方法和规则方法无法完全解决这些难题。

譬如：“باشقىلارنىڭ نەرسىسىنى رۇخسەتسىز ئالما”（未经他人允许不要拿别人的东西），在这句中“ئالما”是个兼类词（动词：不要拿；名词：苹果），在本句中是动词，除此之外，对 “ئالما”进行音节切分后得到两个音节“ما + ئال=ئالما”，其中 “ما”是句子中的表达否定意义的否定后缀，并出现在句末位置。在实验过程中发现，传统统计模型对这句话的分类结果是中性，而本文提出的混合模型使用词性向量、音节向量和位置向量后句子情感正确分类为生气类。

4.3　网络参数对情感分类的影响

本组实验中观察优化函数、dropout、迭代次数、词向量维度等主要网络参数对情感分类的影响。首先，本文模型为了缓解过拟合，使用 dropout 方法进行正则化，通常情况下 dropout 的值选为 0.5，在本次实验中 dropout 的值范围选为 [0.2, 0.3, 0.4, 0.5, 0.6, 0.7]进行实验，其实验结果如图3所示。其次，分别选取 RMSprop、SGD、adadelta、Adam 作为优化函数观察每个优化函数对情感分类的影响，当优化函数选为 Adam 函数时，模型达到了最好的分类效果，其实验结果如图 4 所示。然后，对模型训练迭代次数进行实验，当迭代次数为 19 时达到了最好的分类效果，其结果如图 5 所示。最后对词向量的维度进行实验，本文选取 100、200、300、400 作为词向量的维度观察其对分类效果的影响，当维度为 300 时达到了最好的分类效果，如图 6 所示。

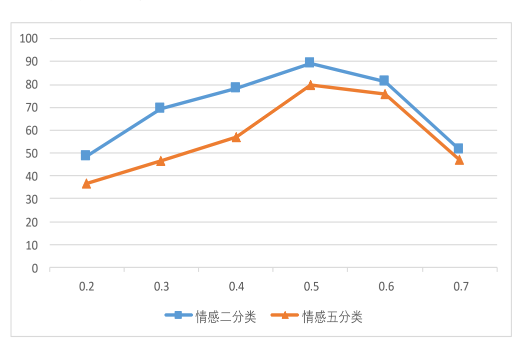


图 3　dropout 参数实验结果

Fig. 3　Dropout parameter experiment result

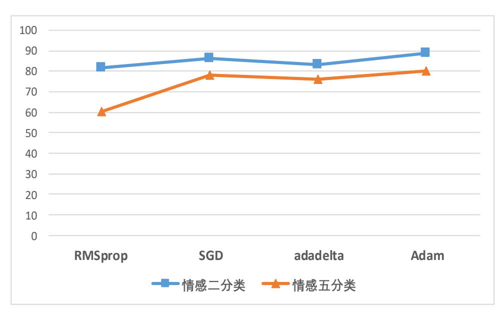


图 4　优化函数实验结果

Fig.4　Optimization function experiment result

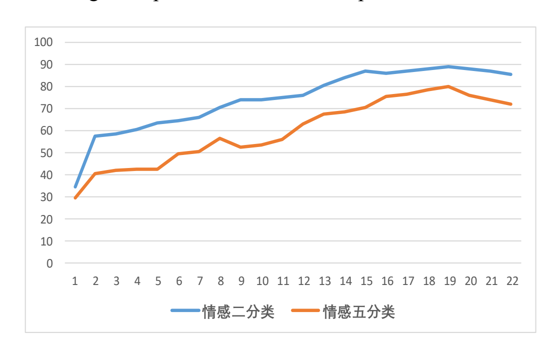


图 5　优化函数实验结果

Fig.5　Experimental results of training epoches

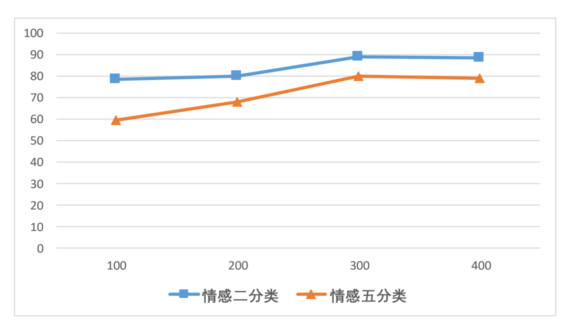


图 6 优化函数实验结果

Fig.6 Experimental results of word vector dimensions

5 结束语

本文针对维吾尔文情感分类任务，提出了一种多特征与带注意力机制的深度学习方法融合的混合模型。首先将词性特征、音节特征和位置特征向量作为词向量的补充，有效地将维吾尔文句子映射到低维抽象特征矩阵，并使用拼接生成的混合向量挖掘单词本身固有的特性；然后通过 BiLSTM 网络对文本进行编码，获取了文本历史上下文依赖信息；随后通过注意力机制更加关注文本中的带情感信息的词语，并将情感词向量与注意力层的输出的隐藏状态向量的拼接向量作为 CNN 网络的输入，从而获取了文本带有情感信息的局部信息；最后使用 *softmax* 函数得到了情感分类结果。实验结果表明，本文提出的混合模型在维吾尔文情感二分类和五分类任务的准确率、召回率和 F1 值上要明显高于传统机器学习方法和简单深度学习方法，本文模型的有效性得到了验证。在今后的工作中，本文将本文模型应用到其他 NLP 任务中评估其效果，并进一步对其网络参数进行优化和改进。

参考文献：

[1] Shin B, Timothy L, Jinho C. Lexicon integrated CNN models with attention for sentiment analysis[J]. arXiv preprint arXiv: 1610. 06272, 2016.

[2] 赵妍妍，秦兵，刘挺. 文本情感分析 [J]. 软件学报. 2010, 21 (8): 1834-1848. (Zhao Yanyan, Qin Bing, Liu Ting. Sentiment analysis [J]. Journal of Software. 2010, 21 (8): 1834-1848.)

[3] Pang B, Lee L, Vaithyanathan S. Thumbs up?: Sentiment classification using machine learning techniques [C]// Proc of ACL-02 Conference on Empirical Methods in Natural Language Processing. [S.l.]:Association for Computational Linguistics, 2002: 79-86.

[4] Wang Yequan, Huang Minlie, Zhu Xiaoyan, *et al.* Attention-based LSTM for aspect-level sentiment classification [C]// Proc of Conference on Empirical Methods in Natural Language Processing. 2016: 606-615.

[5] Tong Yixuan, Zhang Yangsen, Jiang Yuru. Study of sentiment classification for Chinese microblog based on recurrent neural network [J]. Chinese Journal of Electronics. 2016, 25 (4): 601-607.

[6] 刘金硕，李哲，叶馨，等. 文本情感倾向性分析方法：bfsmPMI-SVM. 武汉大学学报：理学版 [J]. 2017, 63 (3): 259-264. (Liu Jishuo, Li Zhe, Ye Xin, *et al.* Sentiment orientation of text: bfsmPMI-SVM[J]. Journal of Wuhan University: Natural Science Edition [J]. 2017, 63 (3): 259-264.)

[7] Zhang Dongwen, Xu Hua, Su Zengcai, *et al.* Chinese comments sentiment classification based on word2vec and SVM perf [J]. Expert Systems with Applications. 2015, 42 (2): 1857–1863.

[8] Wang Jin, Yu Liangchih, Robert Lai K, *et al.* Dimensional sentiment analysis using a regional CNN-LSTM model [C]// Proc of the 54th Annual Meeting of the Association for Computational Linguistics. 2016: 225-230.

[9] Yoon Kim. Convolutional neural networks for sentence classification [C]// Proc of Conference on Empirical Methods in Natural Language Processing. 2014: 1746-1751.

[10] Gichag L, Jeong J, Seungwan S, *et al.* Sentiment classification with word localization based on weakly supervised learning with a convolutional neural network [J]. Knowledge-Based Systems, 2018, 152(7): 70-82.

[11] Hochreiter S, Schmidhuber J. Long short-term memory [J]. Neural Computation. 1997, 9 (8): 1735-1780.

[12] Miyazaki R, Komachi M. Japanese sentiment classification using a tree-structured long short-term memory with attention [Z]. 2017.

[13] Kuta M, Morawiec M, Kitowski J. Sentiment analysis with tree-structured gated recurrent units [Z]. 2017.

[14] Mnih V., Heess N., Graves A., *et al.* Recurrent models of visual attention [C]// Advances in Neural Information Processing Systems, 2014: 2204-2212.

[15] Bahdanau D, Cho K, Bengio Y. Neural machine translation by jointly learning to align and translate [J]. Computer Science, 2014.

[16] Tang Duyu, Qin Bing, Liu Ting. Document modeling with gated recurrent neural network for sentiment classification [C]// Proc of Conference on Empirical Methods in Natural Language Processing. 2015: 1422-1432

[17] Luong M T, Pham H, Manning C D. Effective approaches to attention-based neural machine translation [J]. Computer Science, 2015: 1412-1421

[18] Wu C, Wu F, Liu J, *et al.* THU_NGN at SemEval-2018 task 1: fine-grained tweet sentiment intensity analysis with attention CNN-LSTM[C] //Proc of the 12th International Workshop on Semantic Evaluation. 2018: 186-192.

[19] Yin Wenpeng, Schütze H, Xiang Bing, *et al.* ABCNN: attention-based convolutional neural network for modeling sentence Pairs [J]. Transactions of the Association for Computational Linguistics, arXiv preprint arXiv: 1512. 05193, 2015.

[20] Hochreiter S, Schmidhuber J. Long short-term memory [J]. Neural Computation, 2012, 9 (8): 1735-1780.

[21] Lecun Y, Bengio Y, Hinton G. Deep learning [J]. Nature. 2015, 521 (7553) :436-444.

[22] Mikolov T, Chen Kai, Corrado G, *et al.* Efficient estimation of word representations in vector space [C/OL]// Proc of Workshop at ICLR. 2013. arXiv: 1301. 3781.

[23] Mikolov T, Sutskever I, Chen Kai, *et al.* Distributed representations of words and phrases and their compositionality[C/OL]// Advances in Neural Information Processing Systems,2013, arXiv: 1310. 4546.

[24] Rush A M, Chopra S, Weston J. A neural attention model for sentence summarization [C]//Proc of Conference on Empirical Methods in Natural Language Processing. 2015: 379-389.

[25] Feng Shi, Wang Yang, Liu Liran, *et al.* Attention based hierarchical LSTM network for context-aware microblog sentiment classification [J]. World Wide Web, 2018 (2): 1-23.

[26] Rush A M, Chopra S, Weston J. A neural attention model for sentence summarization [C]// Proc of Conference on Empirical Methods in Natural Language Processing. 2015: 379-389.

[27] Wang Yequan, Huang Minlie, Zhu Xiaoyan, *et al.* Attention-based LSTM for aspect-level sentiment classification [C]// Proc of Conference

on Empirical Methods in Natural Language Processing. 2016: 606-615.
[28] Kalchbrenner N, Grefenstette E, Blunsom P. A convolutional neural network for modelling sentences[C]// Proc of the 52nd Annual Meeting of the Association for Computational Linguistics. 2014: 1749-1751.